Aspects of Localization in Ensemble Kalman Filters

Lars Nerger

Alfred Wegener Institute for Polar and Marine Research
Bremerhaven, Germany

with

P. Kirchgessner, A. Klus, A. Bunse-Gerstner
S. Losa, W. Hiller, J. Schröter, T. Janjic

BremHLR
Kompetenzzentrum für Höchstleistungsrechnen Bremen

ALFRED-WEGENER-INSTITUT
HELMHOLTZ-ZENTRUM FÜR POLAR- UND MEERESFORSCHUNG

University of Reading, July 3, 2014

Outline

Localization – some aspects

- Choosing an optimal localization radius
- Regularizing effect
- Impact on serial observation processing (EnSRF, EAKF)

I will necessarily miss other aspects, e.g.

- Localization and balance
- Adaptive localization
- …

lars.nerger@awi.de

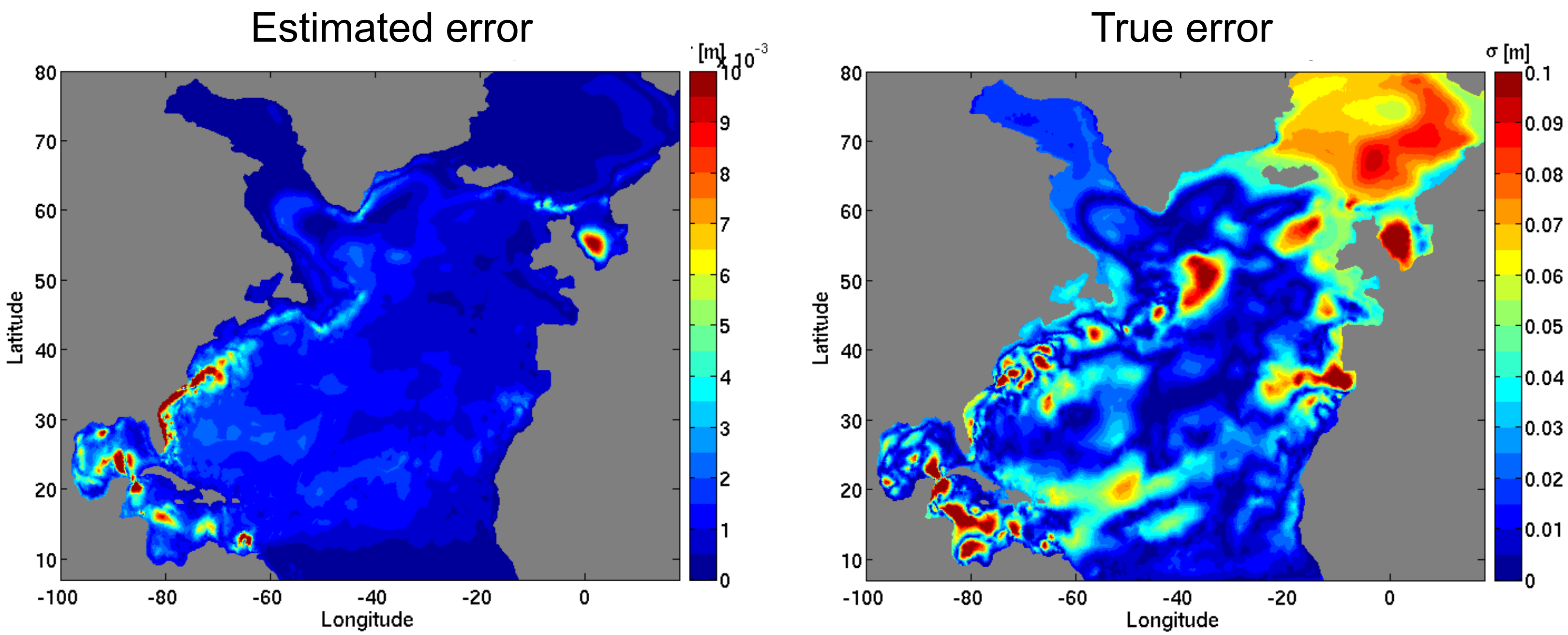
Localization

lars.nerger@awi.de

Motivation for Localization

Ensemble Kalman filter without localization

- Compute globally an optimal combination of state estimate and observations
    - small state corrections
    - strong underestimation of covariances

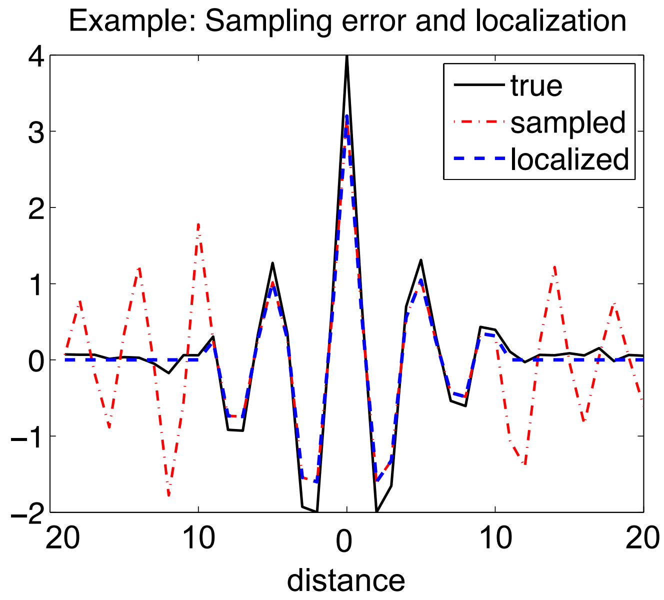


Example: SSH assimilation with SEIK filter and FEOM

lars.nerger@awi.de

Localization: Why and how?

- Combination of observations and model state based on ensemble estimates of error covariance matrices
- Finite ensemble size leads to significant sampling errors
  - errors in variance estimates
  - errors in correlation estimates
    - wrong size if correlation exists
    - spurious correlations when true correlation is zero



- Assume that long-distance correlations in reality are small
  - damp or remove estimated long-range correlations

lars.nerger@awi.de

Effect of Localization

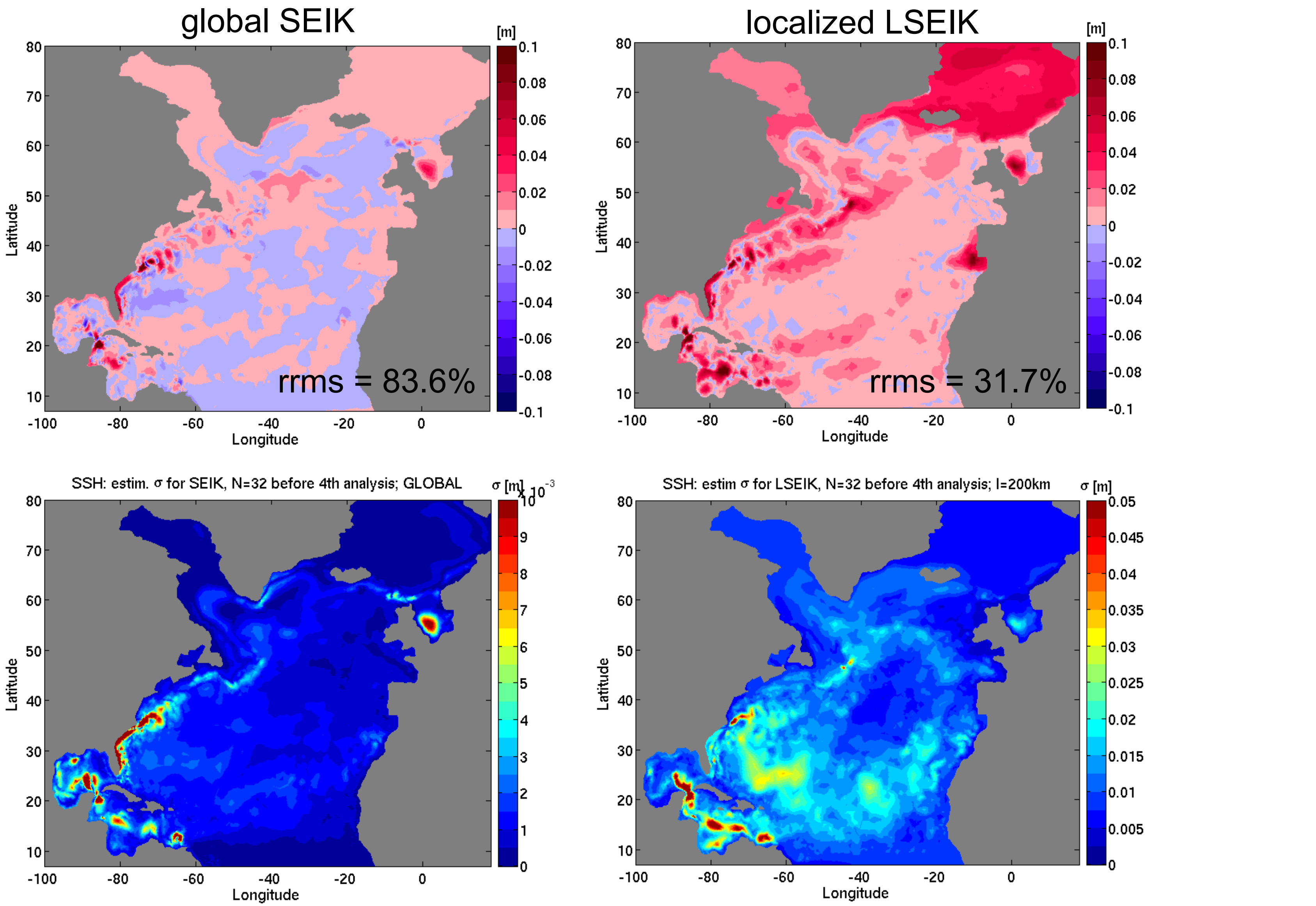
- Remove estimated long-range correlations
  - ➜ Increases degrees of freedom for analysis (globally not locally!)
  - ➜ Increases size of analysis correction
  - ➜ Reduces underestimation of analysis errors

- But:
  - ➜ Also real long-range correlations are removed

Effect of Localization

Stronger correction of states

Large error estimates



lars.nerger@awi.de

Localization Types

Simplified analysis equation:

$$\mathbf{x}^a = \mathbf{x}^f + \frac{\mathbf{P}^f}{\mathbf{P}^f + \mathbf{R}}\left(\mathbf{y} - \mathbf{x}^f\right)$$

Covariance localization

- Modify covariances in forecast covariance matrix $\mathbf{P}^\mathrm{f}$
- Element-wise product with correlation matrix of compact support

Requires that $\mathbf{P}^\mathrm{f}$ is computed (not in ETKF, SEIK, or ESTKF)

E.g.: Houtekamer/Mitchell (1998, 2001), Whitaker/Hamill (2002), Keppenne/Rienecker (2002)

Observation localization

- Modify observation error covariance matrix $\mathbf{R}$
- Needs distance of observation (achieved by local analysis or domain localization)

Possible in all filter formulations

E.g.: Evensen (2003), Ott et al. (2004), Hunt et al. (2007)

lars.nerger@awi.de

Domain & Observation localization

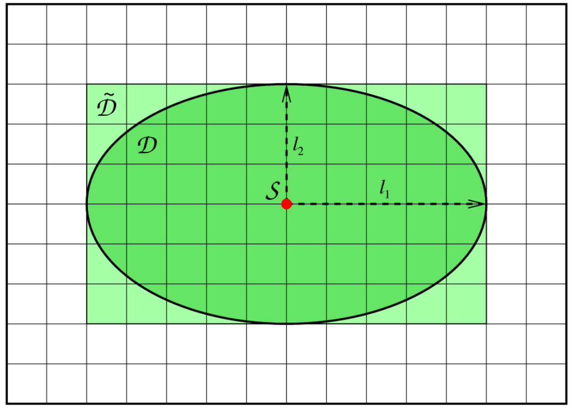
Domain localization

- Perform local filter analysis with observations from surrounding domain

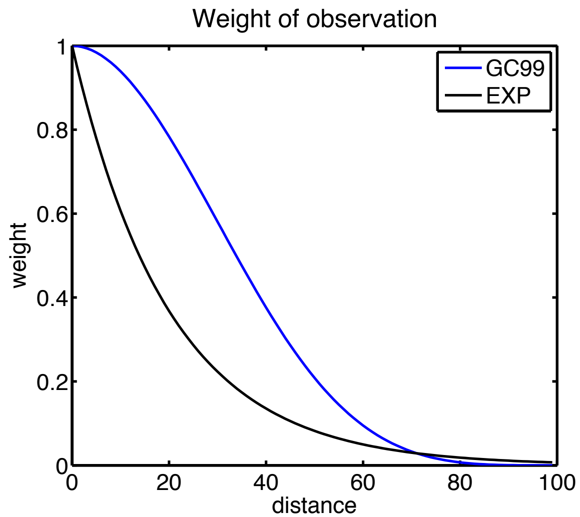
Observation localization

- Use non-unit weight for observations
- reduce weight for remote observations by increasing variance estimate
- use e.g. exponential decrease or polynomial representing correlation function of compact support
- similar, sometimes equivalent, to *covariance localization*

Domain Localization



S: Analysis region
D: Corresponding data region



lars.nerger@awi.de

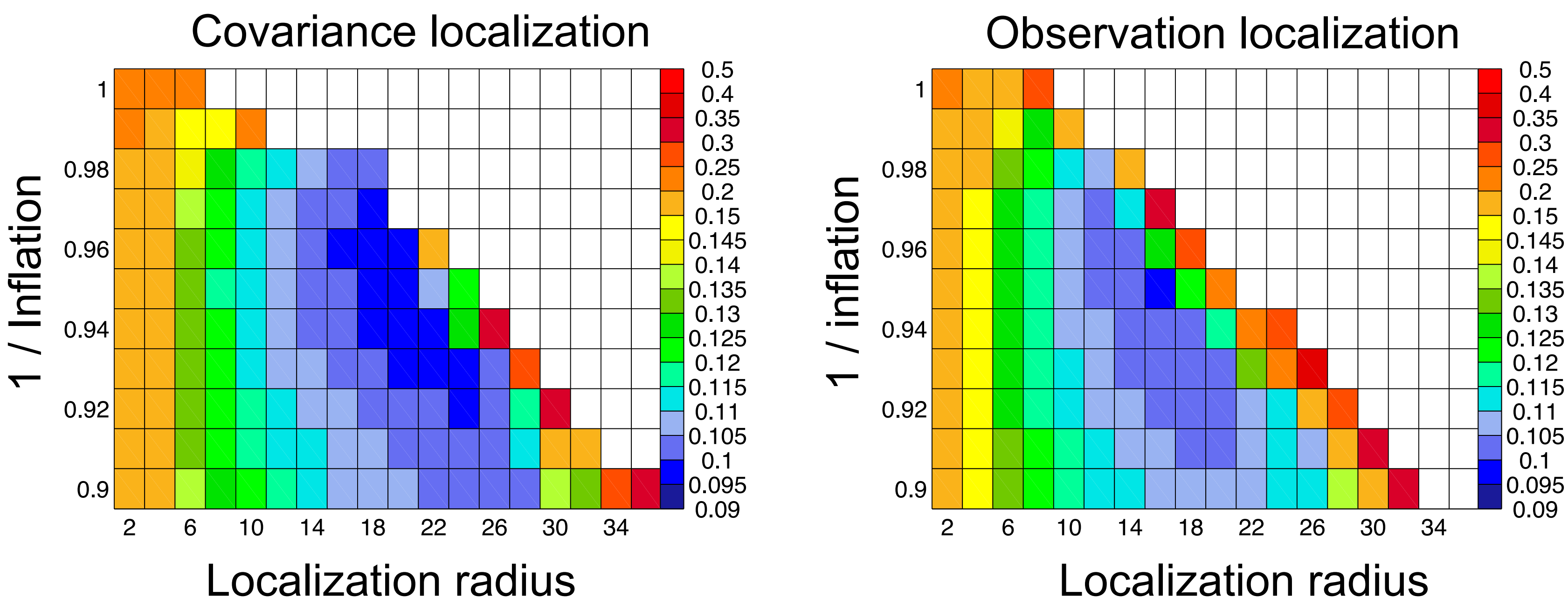
Different effect of localization methods

Experimental result:

- Twin experiment with simple Lorenz96 model
- Use a square-root EnKF and LSEIK
- Covariance localization better than observation localization
  (Also reported by Greybush et al. (2011) with different model)

Time-mean RMS errors

$$\sigma_R = 1.0$$

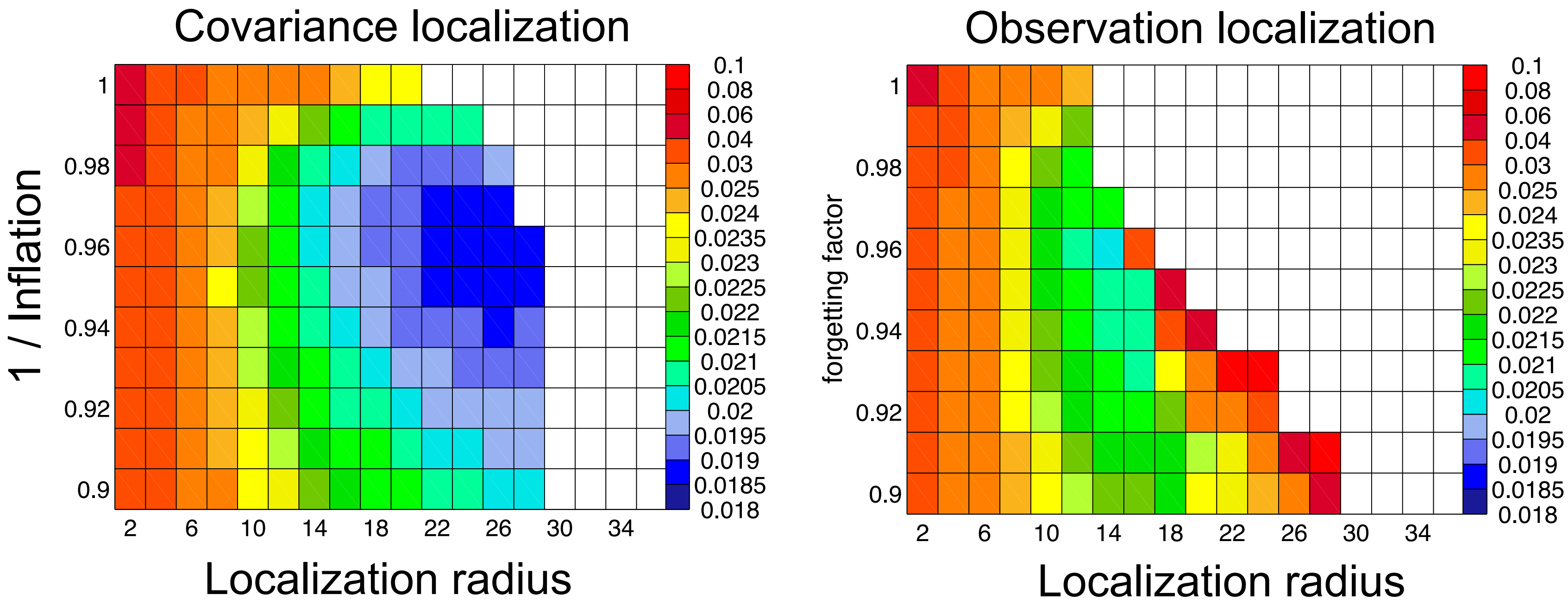


T. Janjic et al., Mon. Wea. Rev. 139 (2011) 2046-2060

Different effect of localization methods (cont.)

Larger differences for smaller observation errors

$$\sigma_R = 0.1$$



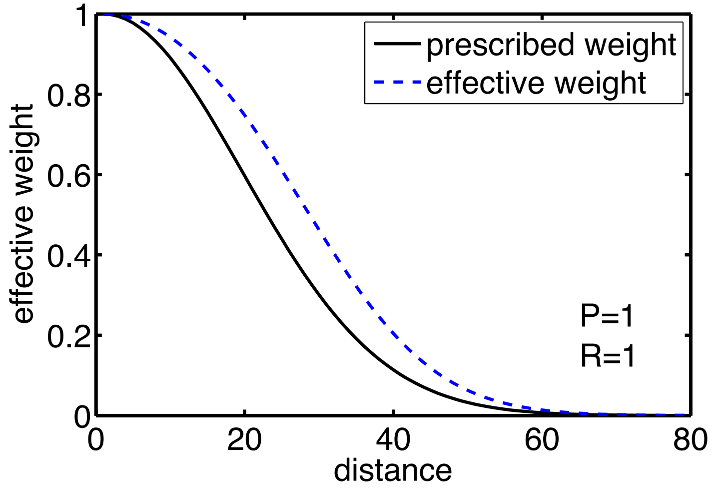
Covariance vs. Observation Localization

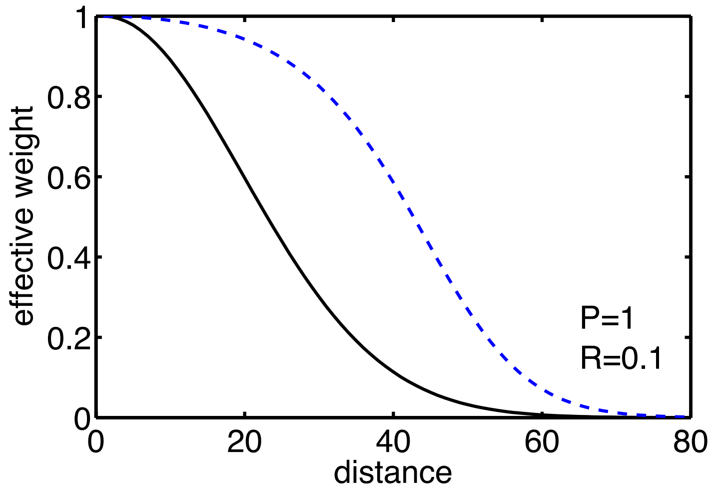
Some published findings:

- Both methods are "similar"
- Slightly smaller width required for observation localization

But note for observation localization:

- Effective localization length depends on errors of state and observations
  - ➢ Small observation error
    ➜ wide localization
  - ➢ Possibly problematic:
    - in initial transient phase of assimilation
    - if large state errors are estimated locally





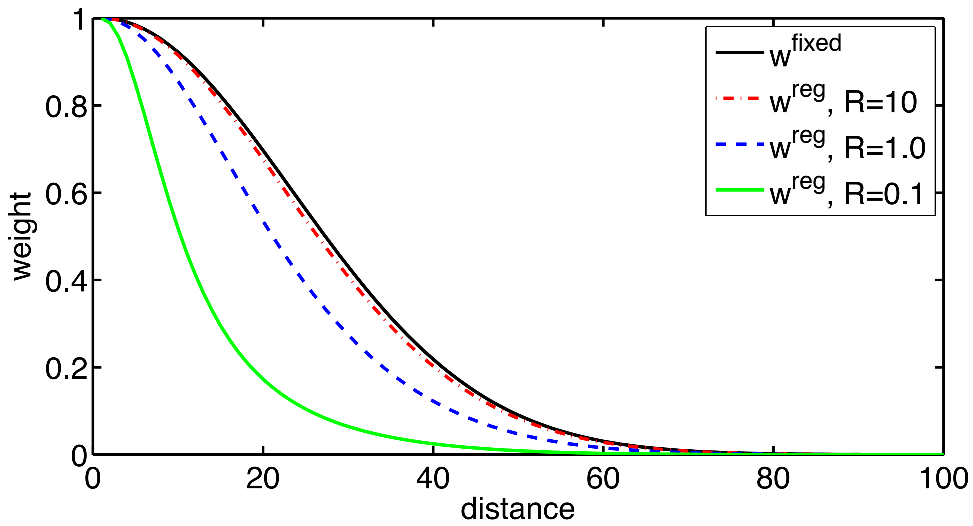
P: state error variance
R: observation error variance

lars.nerger@awi.de

Regulated Localization

➜ New localization function for observation localization

- formulated to keep effective length constant (exact for single observation)
- depends on state and observation errors
- depends on fixed localization function
- cheap to compute for each observation
- Only exact for single observation – works for multiple

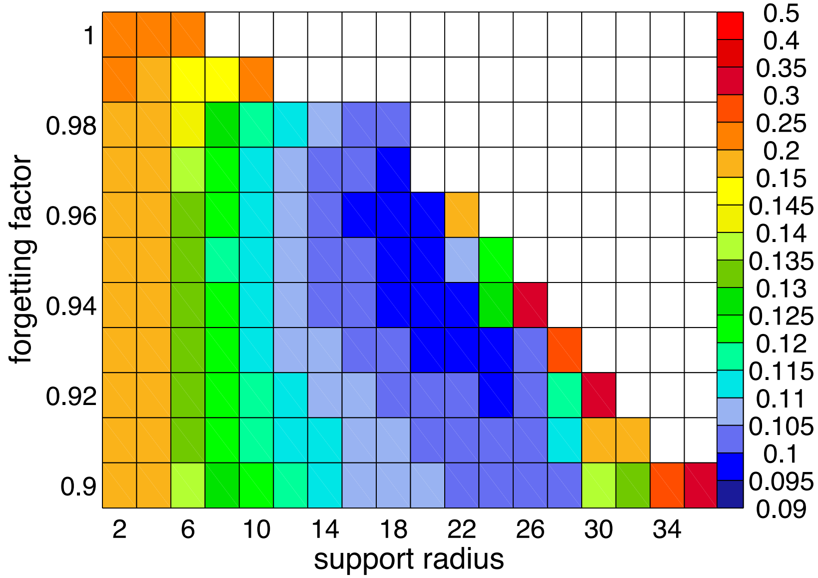


P=1

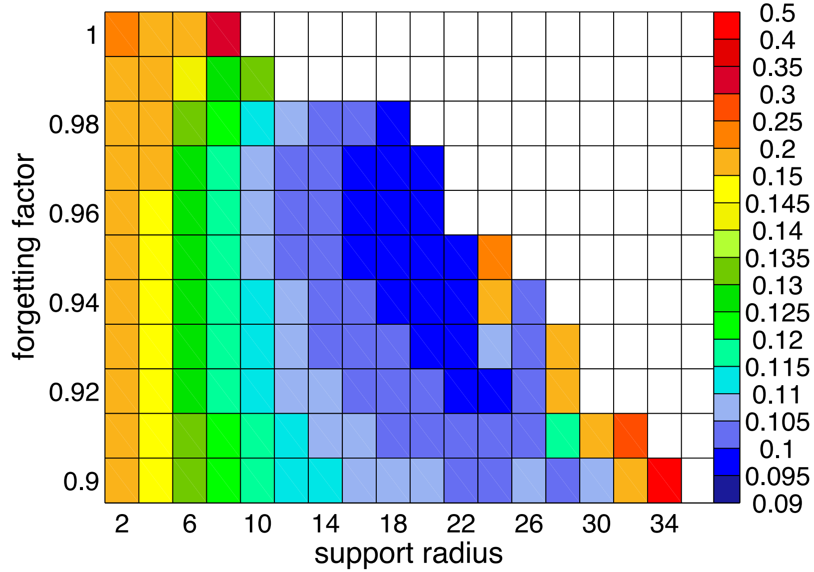
L. Nerger et al. QJ Royal. Meterol. Soc. 138 (2012) 802-812

Lorenz96 Experiment: Regulated Localization

Covariance loc., N=10, R=0.5



Regulated localization, N=10, R=0.5



- Reduced minimum rms errors
- Increased stability region
- Description of effective localization length explains the findings of other studies!
- Impact also with FESOM ocean model (but smaller)

lars.nerger@awi.de

Optimal Localization Radius

(Paul Kirchgessner et al.)

lars.nerger@awi.de

Domain & Observation localization

Localization radius can depend on

- Ensemble size
- Model dynamics & resolution
- Field

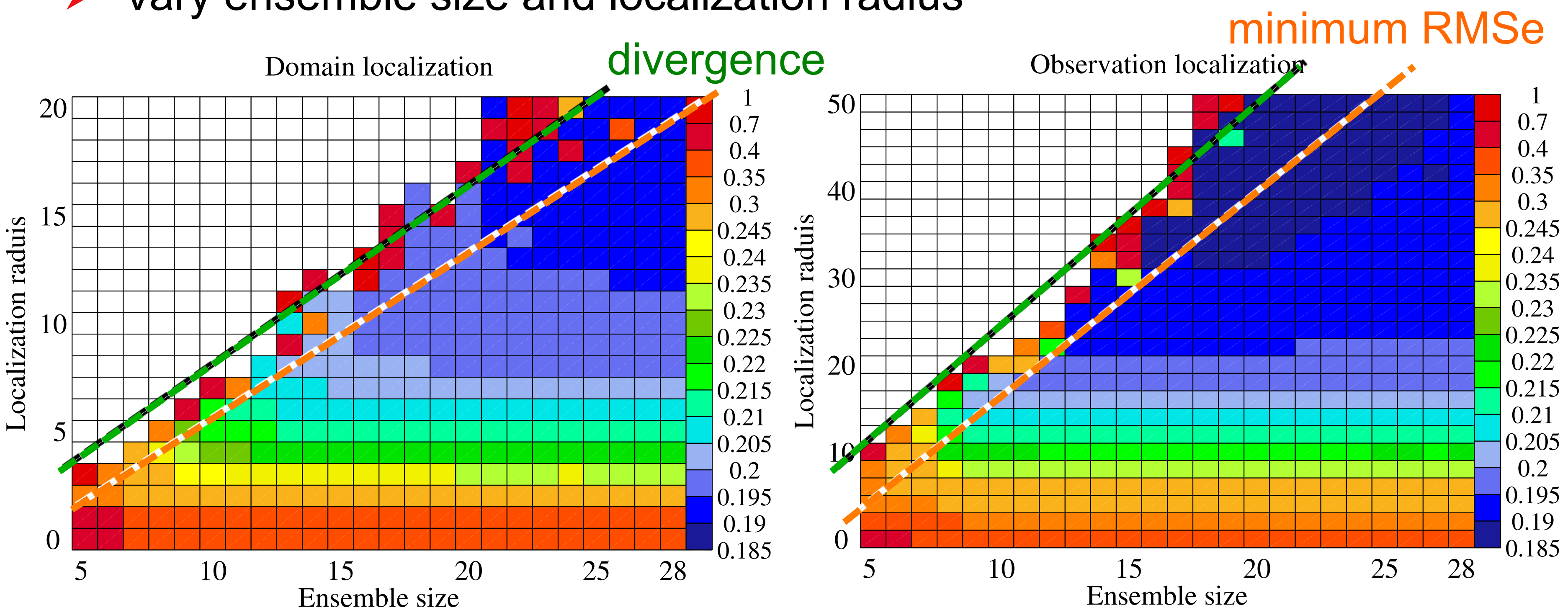
Optimal localization radius

- Typically determined experimentally (very costly)
- Some authors proposed adaptive methods (e.g. Anderson, Bishop/Hodyss)
  - still with tunable parameters

lars.nerger@awi.de

Relation between ensemble size and localization radius

- Test runs with Lorenz-96 model
- Vary ensemble size and localization radius

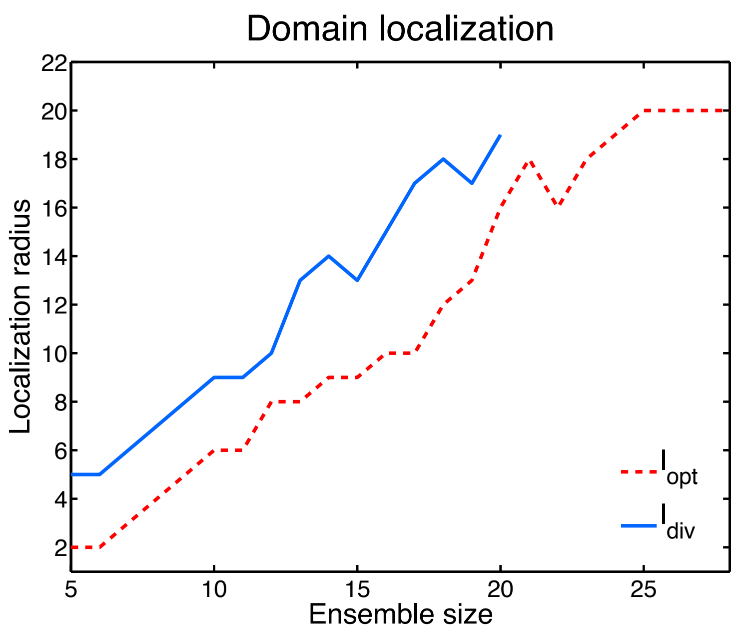


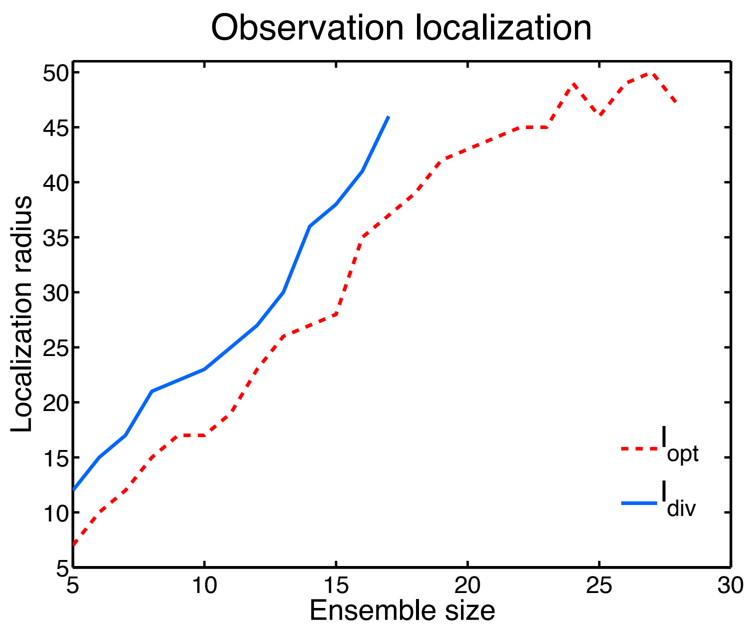
- White: Filter divergence

lars.nerger@awi.de

Optimal localization radius

- Optimal localization radius as function of ensemble size





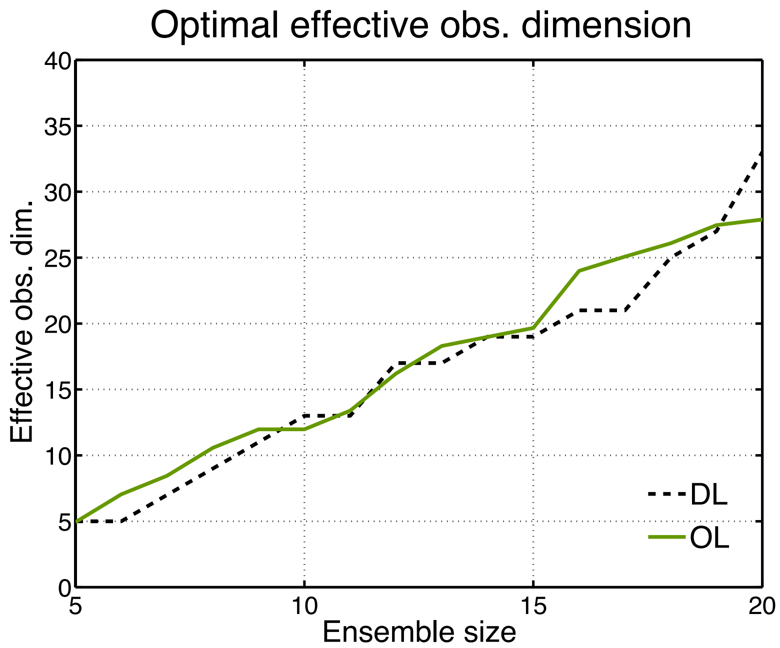
- Linear dependence for domain and observation localization
- Radius larger for OL than DL

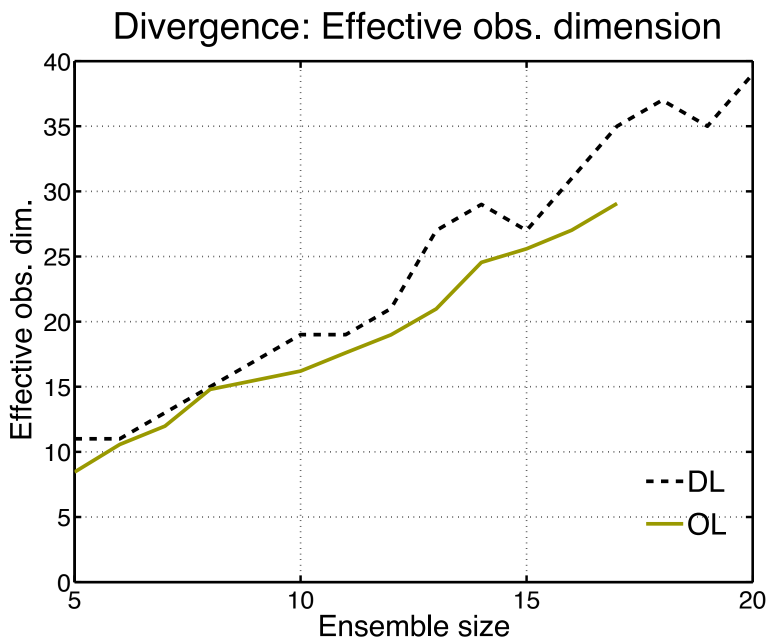
lars.nerger@awi.de

Relate domain and observation localizations

- Define:

  Effective observation dimension d_W = sum of observation weights





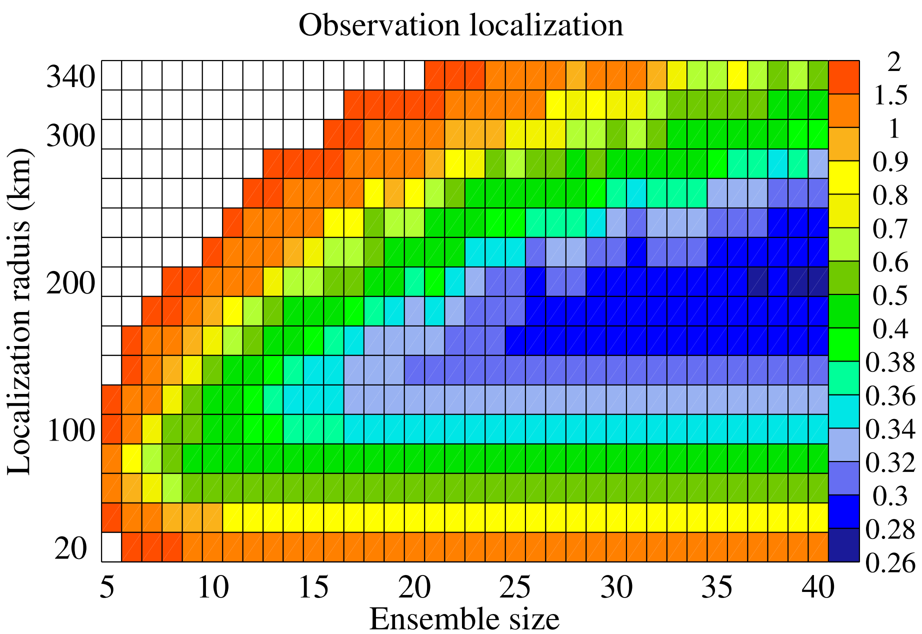
- Minimum RMS errors when effective obs. dimension slightly larger than ensemble size

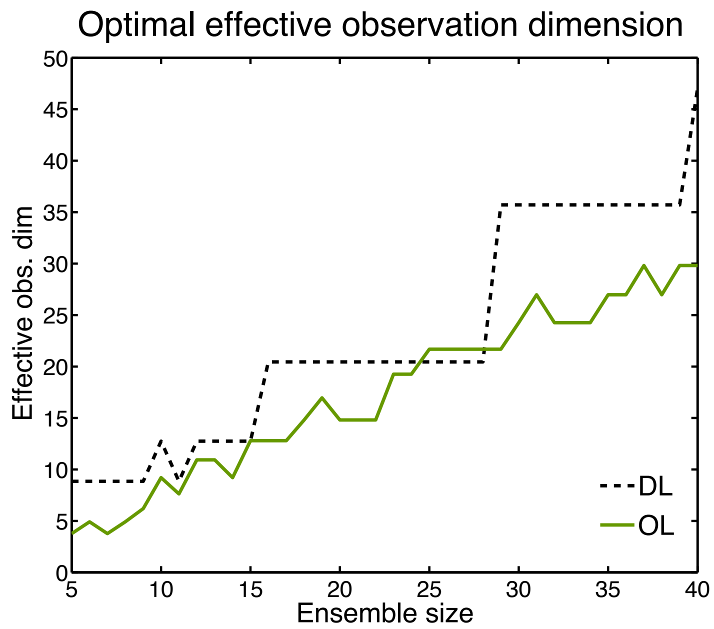
- When d_W=ensemble size, errors are almost as small (optimal use of degrees of freedom from ensemble?)

P. Kirchgessner et al. Mon. Wea. Rev. 142 (2014) 2165-2175

2D Shallow Water Model

- Shallow water model simulating a double gyre in a box
- Assimilate sea surface height at each grid point





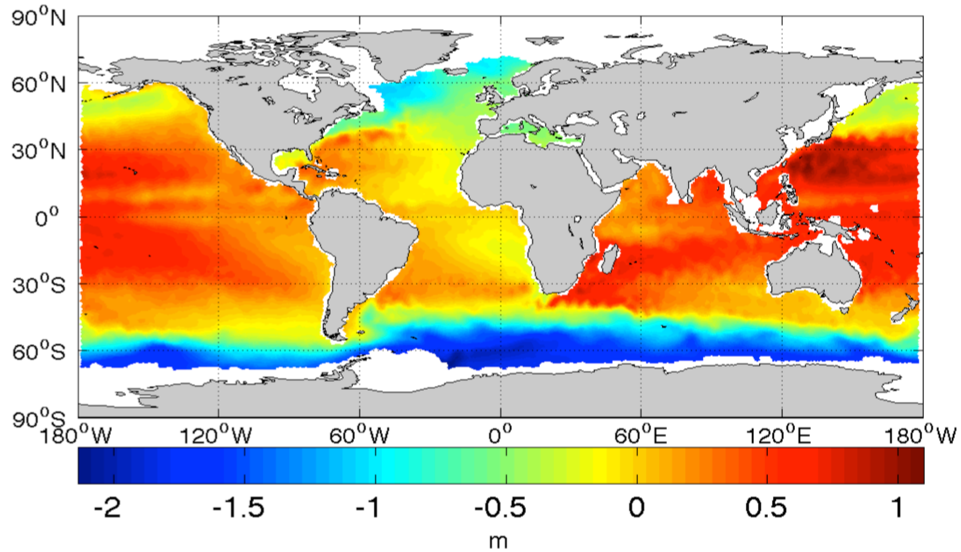
- For DL: steps due to addition of observations
- d_W optimal if about or slightly lower than ensemble size
- relation holds for different weight functions

P. Kirchgessner et al. Mon. Wea. Rev. 142 (2014) 2165-2175

Large scale data assimilation: Global ocean model

- Finite-element sea-ice ocean model (FESOM)
- Global configuration (~1.3 degree resolution with refinement at equator)
- State vector size: 10^7
- Scales well up to 256 processor cores

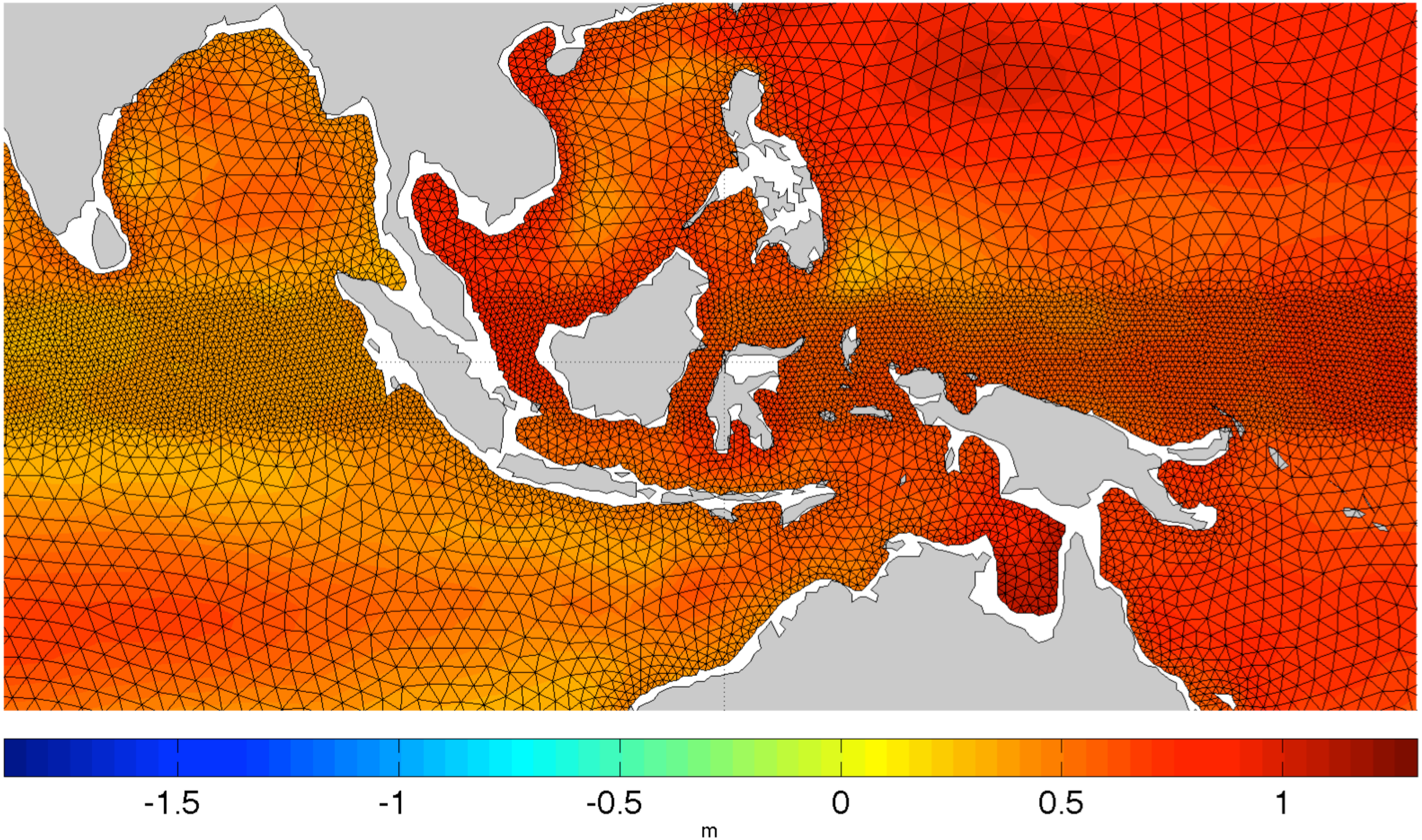
Sea surface elevation



- Assimilate synthetic sea surface height data for ocean state estimation
- Very costly due to large model size (using up to 2048 processor cores)

lars.nerger@awi.de

Model mesh at the equator

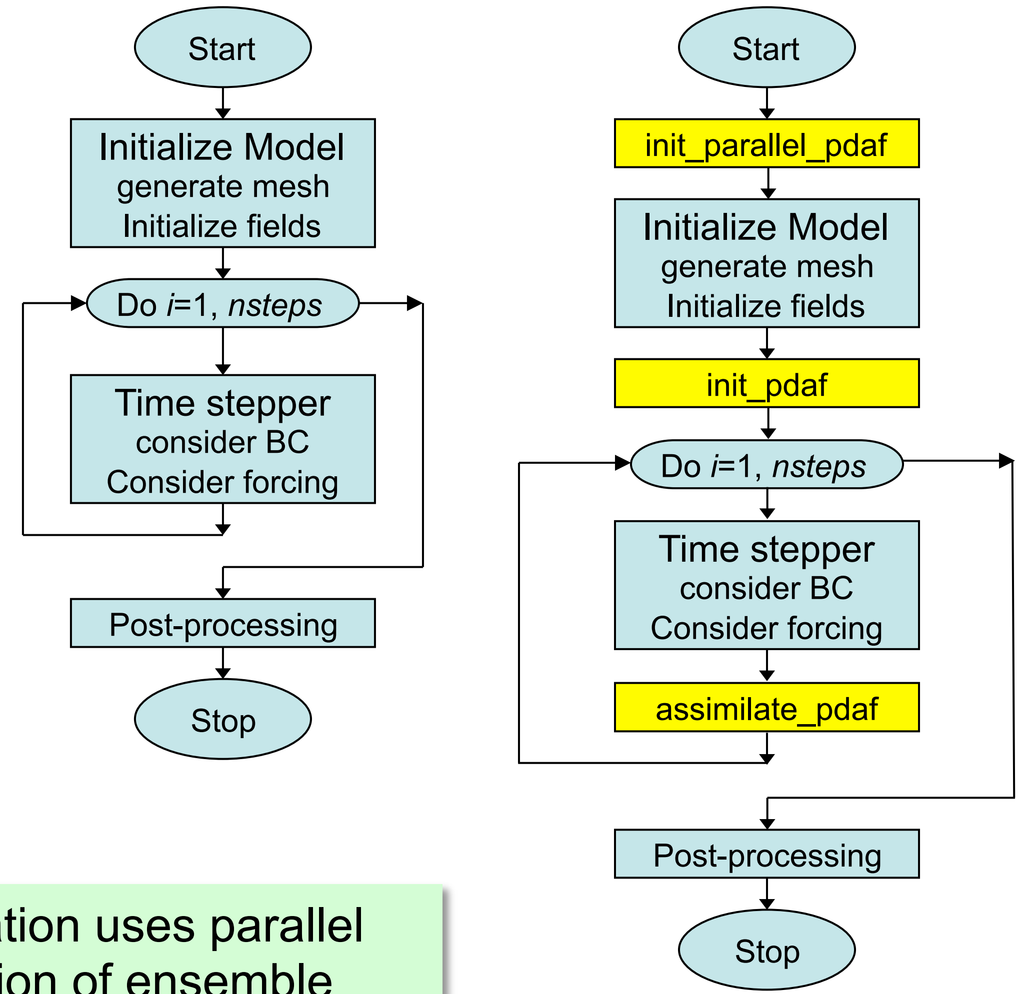


lars.nerger@awi.de

Extending a Model for Data Assimilation

PDAF Parallel Data Assimilation Framework

Model



Extension for data assimilation

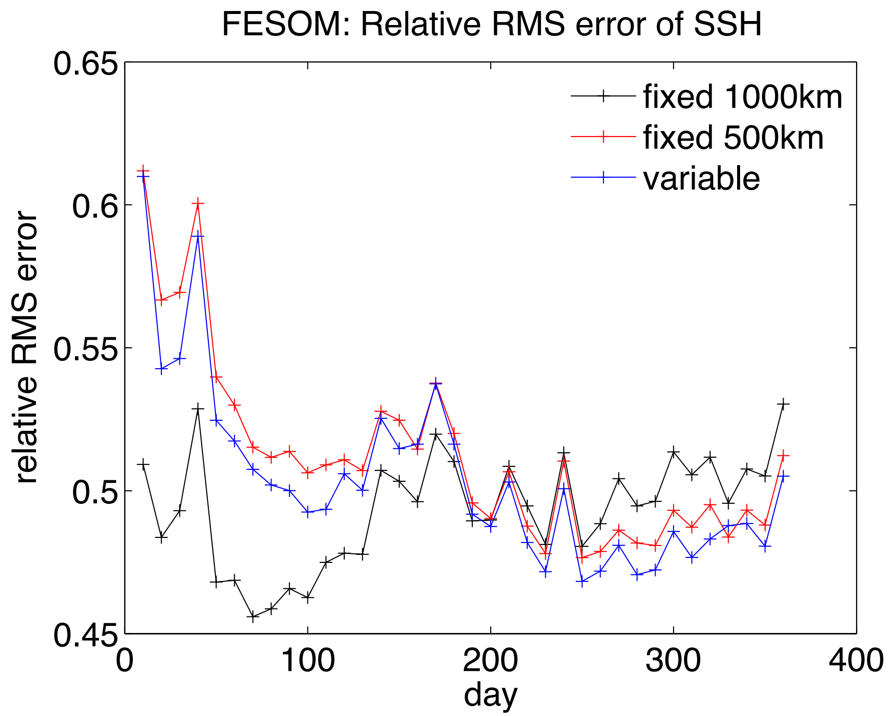
Implementation uses parallel configuration of ensemble forecast provided by PDAF

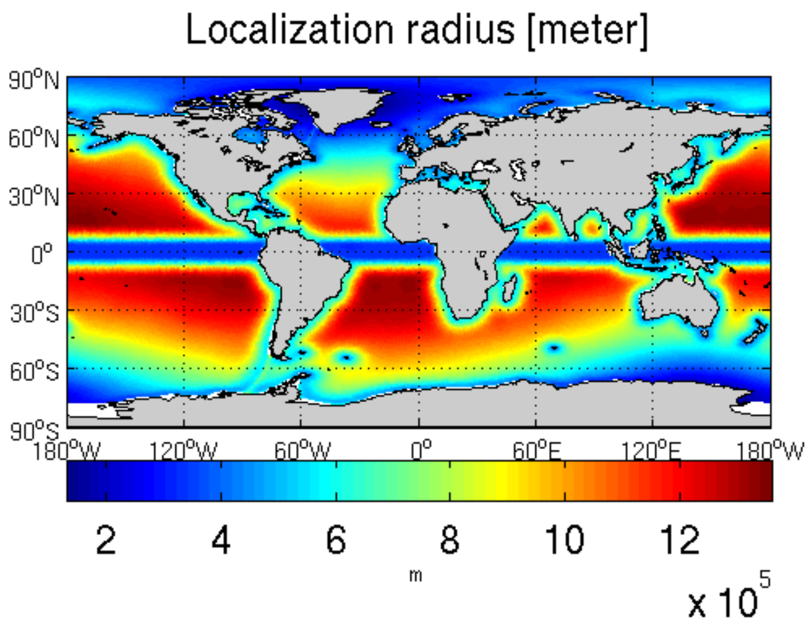
Open source:
Code and documentation available at

http://pdaf.awi.de

Adaptive localization radius in global ocean model

- Localization radius follows mesh resolution
- Fixed 1000km radius leads to increasing errors in 2nd half of year
- Lower RMS error in SSH than fixed 500km radius





lars.nerger@awi.de

Discussion on localization radius

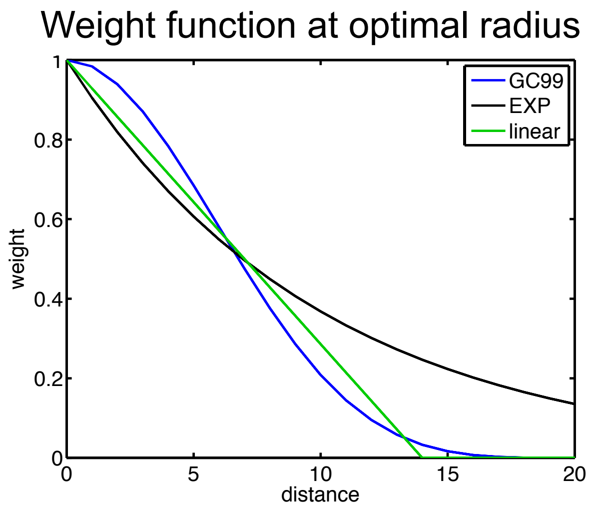
- Findings:
  - Effective observation dimension d_W relates to degrees of freedom
  - d_W close to ensemble size a good choice
  - No dependence on model dynamics

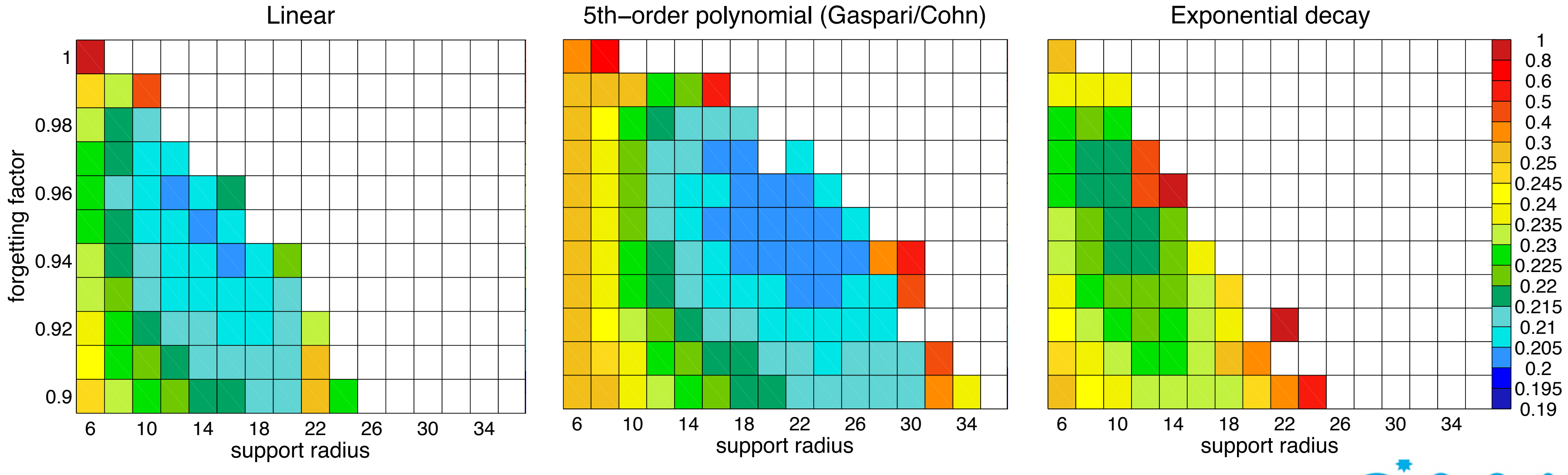
- Limitations
  - Observations at each grid point (optimal d_W smaller for incomplete observations)
  - Uniform observation error
  - Ignoring information content of observations (e.g. Migliorini, QJRMS 2013)

P. Kirchgessner et al. Mon. Wea. Rev. 142 (2014) 2165-2175

Weight function

- Why 5th-order Gaspari/Cohn polynomial?
- Covariance function not required for OL
- Furrer/Bengtsson (2007) indicate best sampling error reduction in $\mathbf{P}^f$ for exponential covariances
- For Lorenz96, some other functions give similar errors – but not significantly lower ones





lars.nerger@awi.de

AWI

Localization

as Regularization

(Master thesis Andrea Klus @U Bremen)

lars.nerger@awi.de

Regularization in Ensemble Kalman Filters

- Write Kalman filter analysis as minimization

$$||(\mathbf{P}^f)^{-1/2}\boldsymbol{\delta}||_2^2 + ||\mathbf{R}^{-1/2}\mathbf{H}\boldsymbol{\delta} - \mathbf{R}^{-1/2}(\mathbf{y} - \mathbf{H}\mathbf{x}^f)||_2^2 = min!$$

  with $\boldsymbol{\delta} = \mathbf{x} - \mathbf{x}^f$

- General form (not the same x, y)

$$||\mathbf{A}\mathbf{x} - \mathbf{y}||_2 + \lambda||\mathbf{L}\mathbf{x}||_2 = min!$$

  (standard Tikhonov regularization for $\mathbf{L}=\mathbf{I}$)

- For **ETKF**

  Use $\boldsymbol{\delta} = \mathbf{V}\boldsymbol{\omega}$ with $\mathbf{V}\mathbf{V}^T = \mathbf{P}^f$

  then

$$||\boldsymbol{\omega}||_2^2 + ||\mathbf{R}^{-1/2}\mathbf{H}\mathbf{V}\boldsymbol{\omega} - \mathbf{R}^{-1/2}(\mathbf{y} - \mathbf{H}\mathbf{x}^f)||_2^2 = min!$$

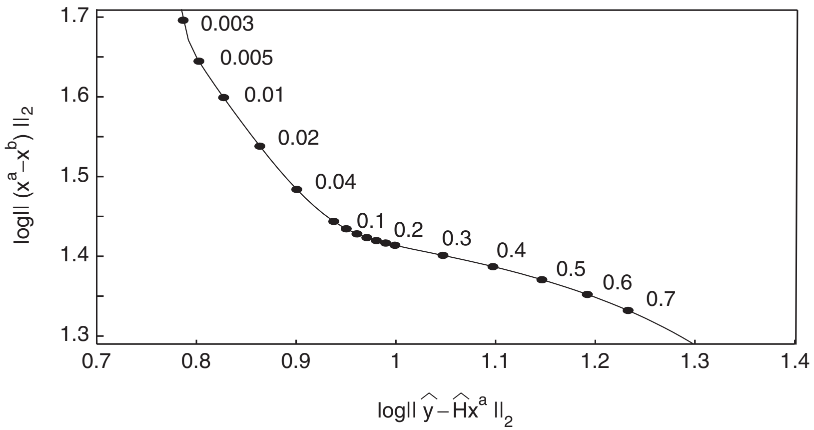
L-curves

- Examine norm of both terms on minimization problem varying λ

$$||\mathbf{Ax} - \mathbf{y}||_2 + \lambda||\mathbf{Lx}||_2 = min!$$

- Plot residual term vs. penalty term

- Example for 4D-Var case (Johnson, Nichols, Hoskins, IJNMF, 2005)



with

$$\tilde{\mathscr{J}}(\chi) = \mu^2 \|\chi\|_2^2 + \|\mathbf{C}_R^{-1/2}\hat{\mathbf{d}} - \mathbf{C}_R^{-1/2}\hat{\mathbf{H}}C_B^{1/2}\chi\|_2^2$$

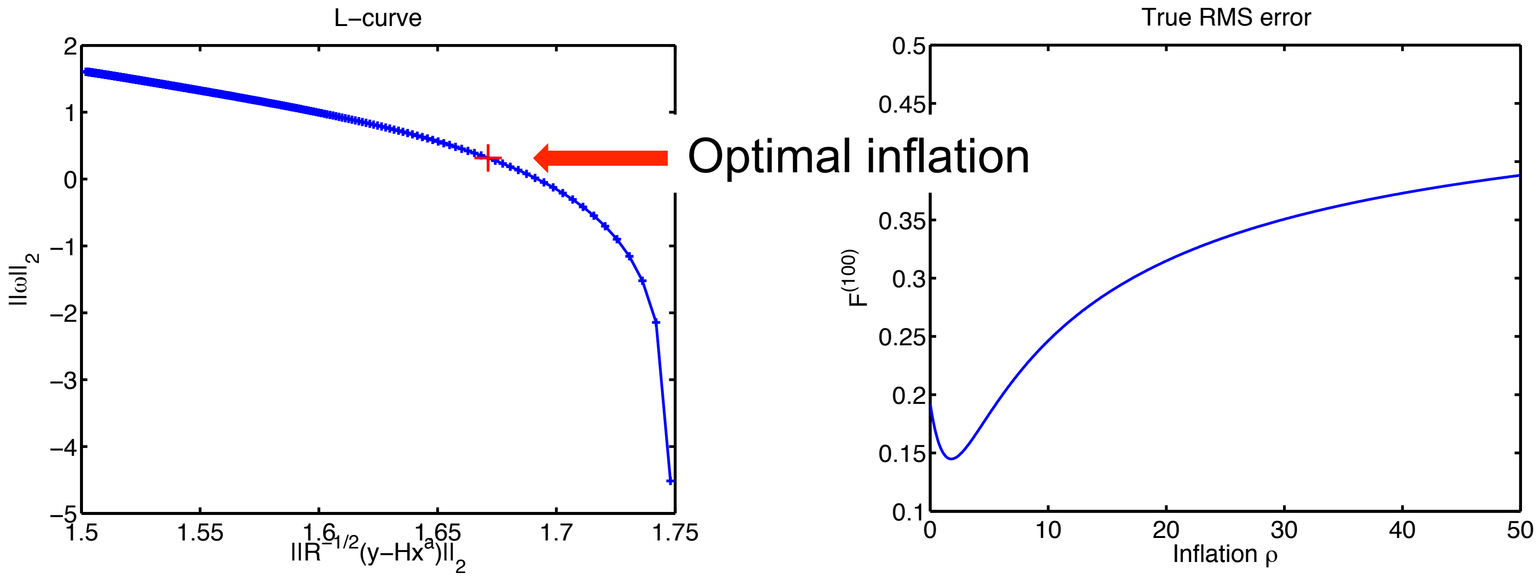
lars.nerger@awi.de

ETKF with Inflation

- Inflation is a standard method to stabilize ensemble filters
- Modify minimization problem to

$$|\rho^{-1}| \cdot ||\boldsymbol{\omega}||_2^2 + ||\mathbf{R}^{-1/2}\mathbf{H}\mathbf{V}\boldsymbol{\omega} - \mathbf{R}^{-1/2}(\mathbf{y} - \mathbf{H}\mathbf{x}^f)||_2^2 = min!$$

- L-curve (for Lorenz-96 at time step 100 and spin up with ρ=1.05)



(not an ‚L‘)

lars.nerger@awi.de

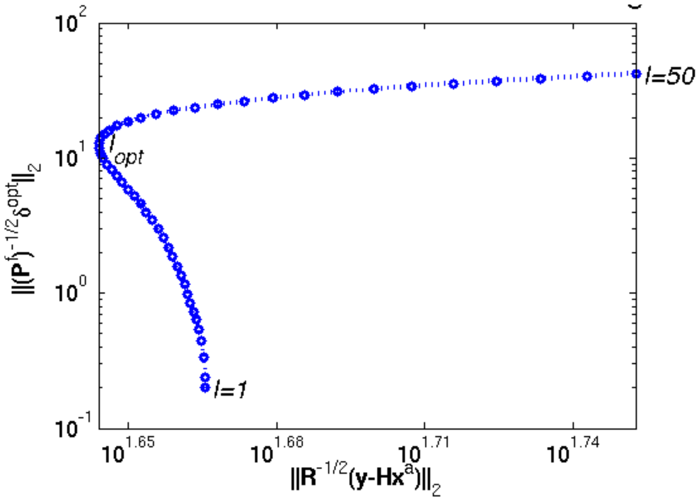
EnKF with covariance localization

- Minimize

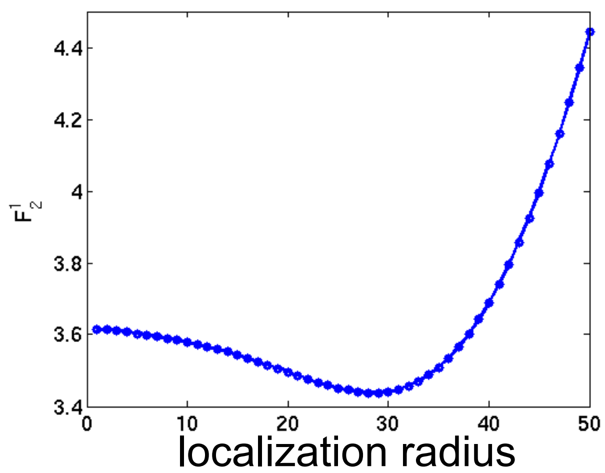
$$||(\mathbf{C} \circ \mathbf{P}^f)^{-1/2}\boldsymbol{\delta}||_2^2 + ||\mathbf{R}^{-1/2}\mathbf{H}\boldsymbol{\delta} - \mathbf{R}^{-1/2}(\mathbf{y} - \mathbf{H}\mathbf{x}^f)||_2^2 = min!$$

- Vary localization radius l defining $\mathbf{C}$
- Use 5th order polynomial for $\mathbf{C}$ (Gaspari/Cohn, 1999)

L-curve



Analysis error variance



lars.nerger@awi.de

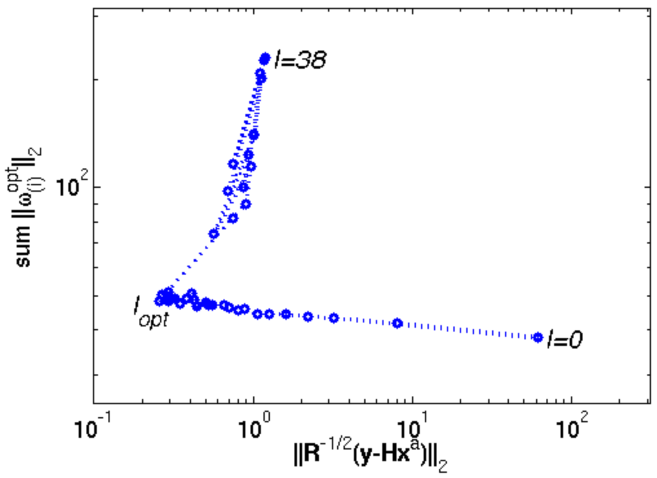
ETKF with observation localization

- Minimize the local problem

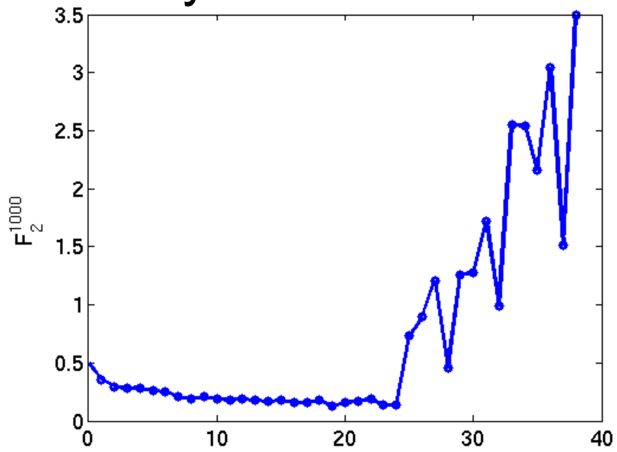
$$||(\tilde{\mathbf{C}}_{loc(i)} \circ (\mathbf{T}\mathbf{R}^{-1}\mathbf{T}_i^T))^{1/2}(\mathbf{T}_i\mathbf{H}\mathbf{V})\boldsymbol{\omega}_{loc(i)} - (\tilde{\mathbf{C}}_{loc(i)} \circ (\mathbf{T}\mathbf{R}^{-1}\mathbf{T}_i^T))^{1/2}\mathbf{T}_i(\mathbf{y} - \mathbf{H}\mathbf{x}^f)||_2^2 + ||\boldsymbol{\omega}_{loc(i)}||_2^2 = min!$$

- Vary localization radius l defining $\mathbf{C}$ and $\mathbf{T}$
- Consider minimization at a single grid point, sum over all points

L-curve



Analysis error variance



localization radius

(similar behavior at all grid points)

lars.nerger@awi.de

Discussion on regularization

- Localization regularizes the filter analysis
- Analysis for optimal radius is a posteriori
- Can we utilize it in practice?

lars.nerger@awi.de

Impact of localization

on serial observation processing

(EnSRF, EAKF)

lars.nerger@awi.de

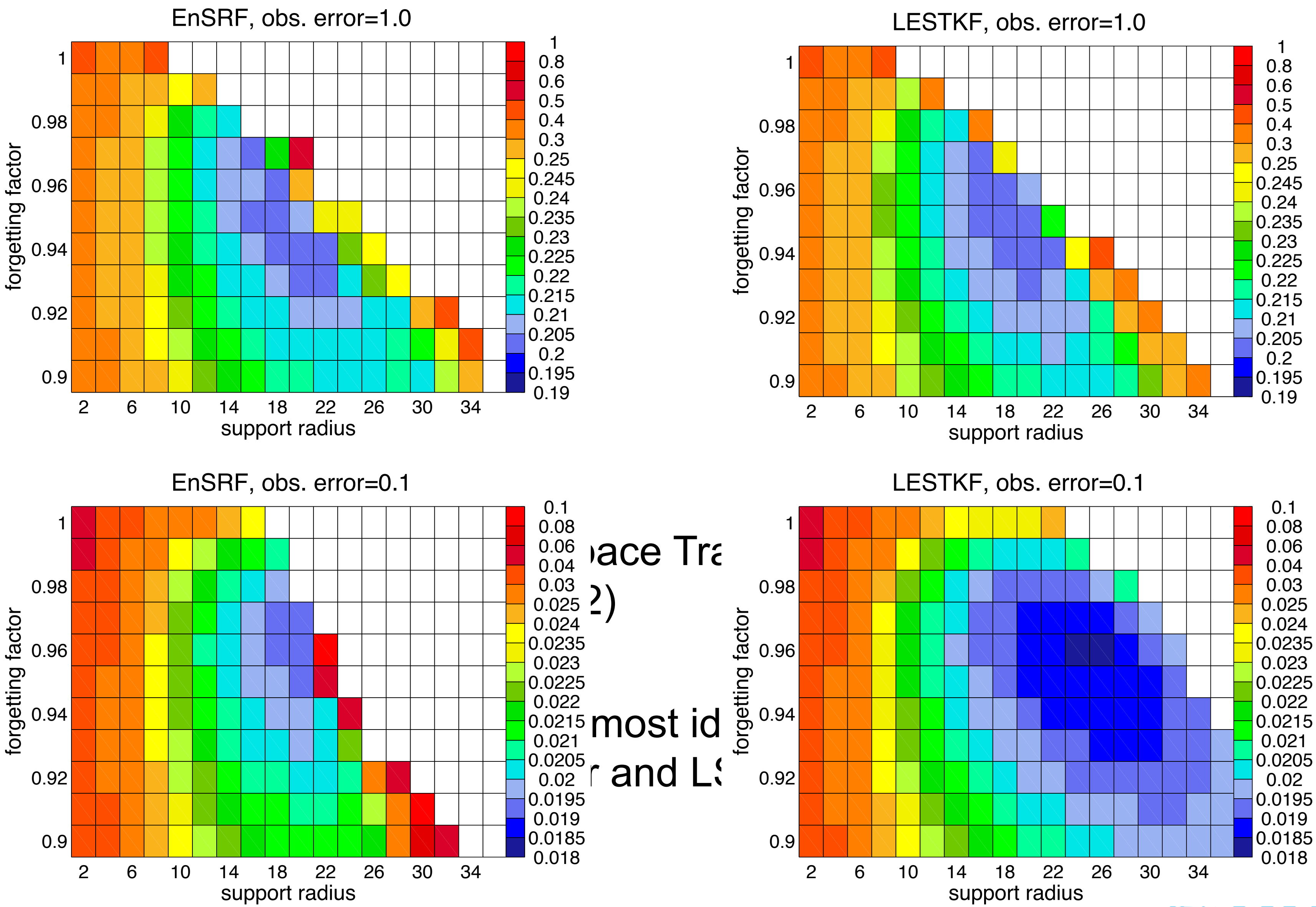
Serial observation processing

Synchronous assimilation	Serial observation processing
ETKF, SEIK, ESTKF, (EnKF)	***EnSRF, EAKF***
• Assimilation all observation at a given time at once	• Perform a loop assimilating each single observation
• Usually using ensemble-space transformations	• Efficient: Avoids matrix-matrix operations
• Possible for arbitrary observation error covar. matrices	• Requires diagonal observation error covar. matrix

Use **observation localization**	Use **covariance localization**

lars.nerger@awi.de

Test with Lorenz96



ace Tra
2)

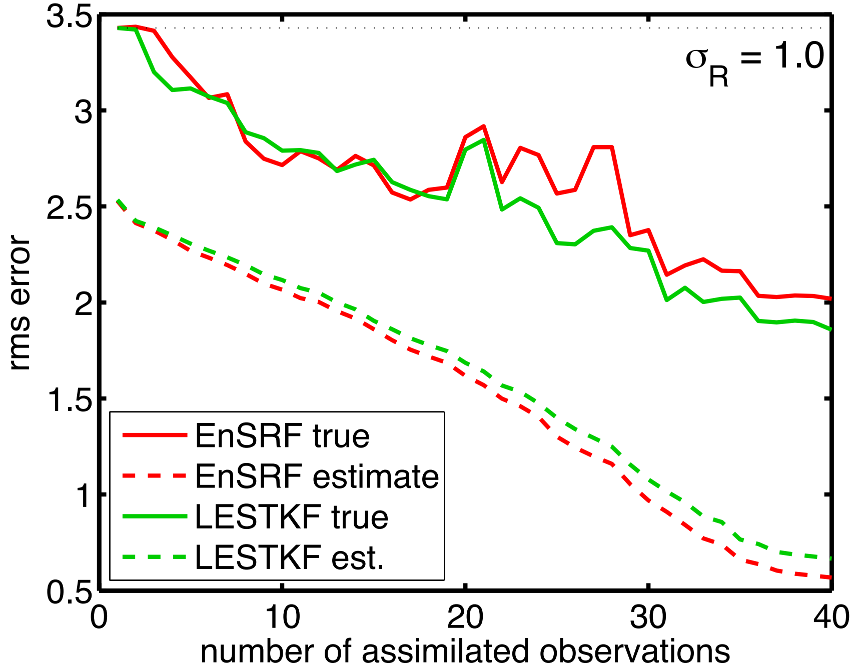
most id
r and L

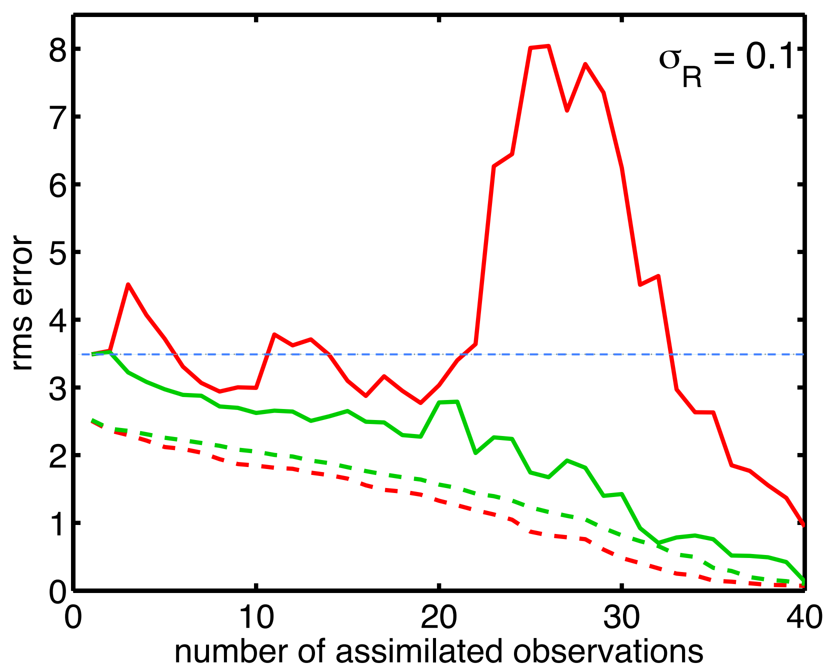
RMS error over number of observations

How does the RMS error develop during the loop over all observations?

At first analysis step:

- EnSRF: Compute RMS errors at each iteration
- LESTKF: Do 40 experiments with increasing number of obs.



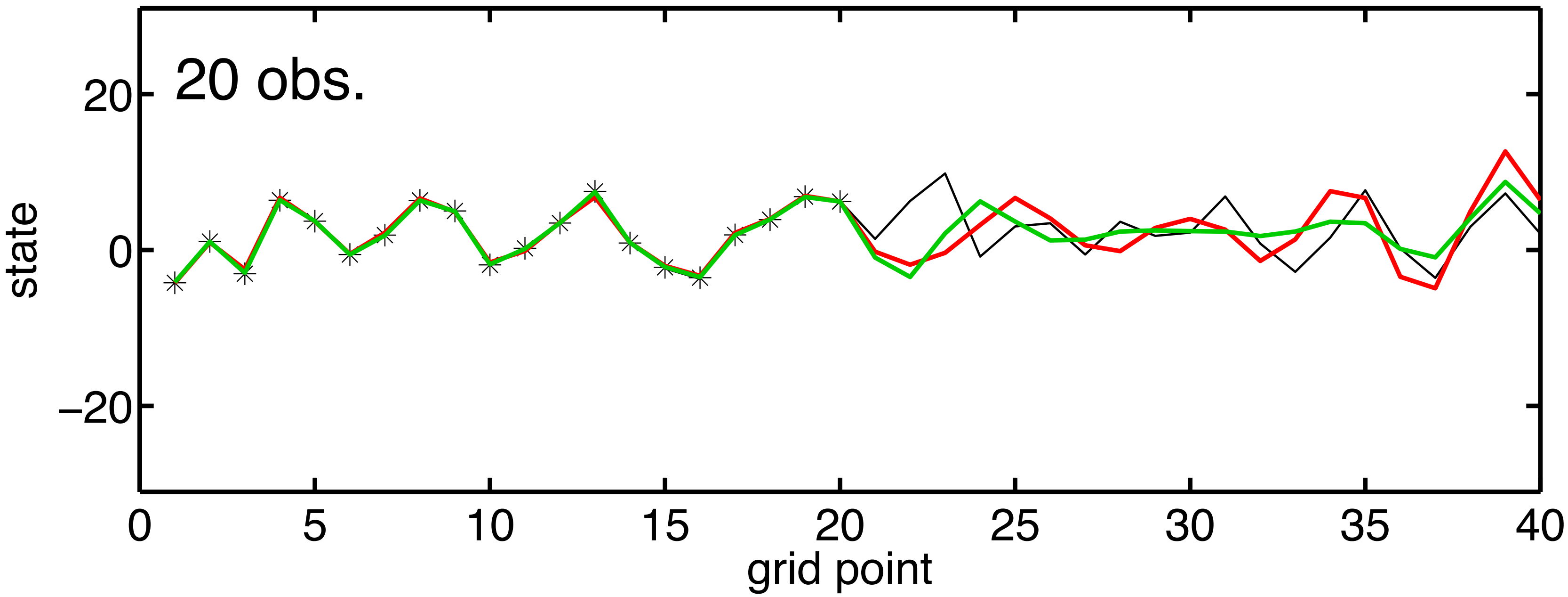


lars.nerger@awi.de

Instability of serial obs. Processing with localization

More detailed view:

- State estimate for different numbers of observations



lars.nerger@awi.de

Inconsistent matrix updates

The Kalman filter updates the covariance matrix according to

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\,\mathbf{P}^f\,(\mathbf{I} - \mathbf{KH})^T + \mathbf{KRK}^T \qquad (1)$$

With the Kalman gain

$$\mathbf{K} = \mathbf{P}^f\mathbf{H}^T\left(\mathbf{HP}^f\mathbf{H}^T + \mathbf{R}\right)^{-1} \qquad (2)$$

this simplifies to

$$\mathbf{P}^a = (\mathbf{I} - \mathbf{KH})\,\mathbf{P}^f \qquad (3)$$

(1) and (3) yield same result **only** with gain (2)!

Not fulfilled with localization:

$$\mathbf{K}_{loc} = \left(\mathbf{C} \circ \mathbf{P}^f\right)\mathbf{H}^T\left(\mathbf{H}\left(\mathbf{C} \circ \mathbf{P}^f\right)\mathbf{H}^T + \mathbf{R}\right)^{-1}$$

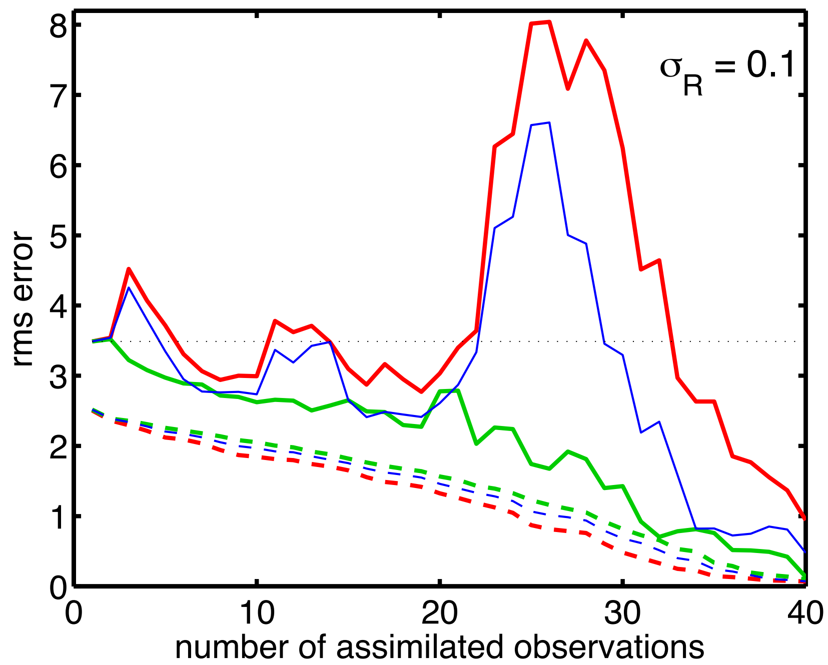
- Update of **P** is inconsistent in localized EnSRF (already noted by Whitaker & Hamill (2002), but never further examined)

Inconsistent matrix updates (2)

The inconsistency also occurs in LETKF, LESTKF, LSEIK, EnKF ...

- But here: update is only done once followed by ensemble forecast

- LESTKF with serial observation processing also shows instability



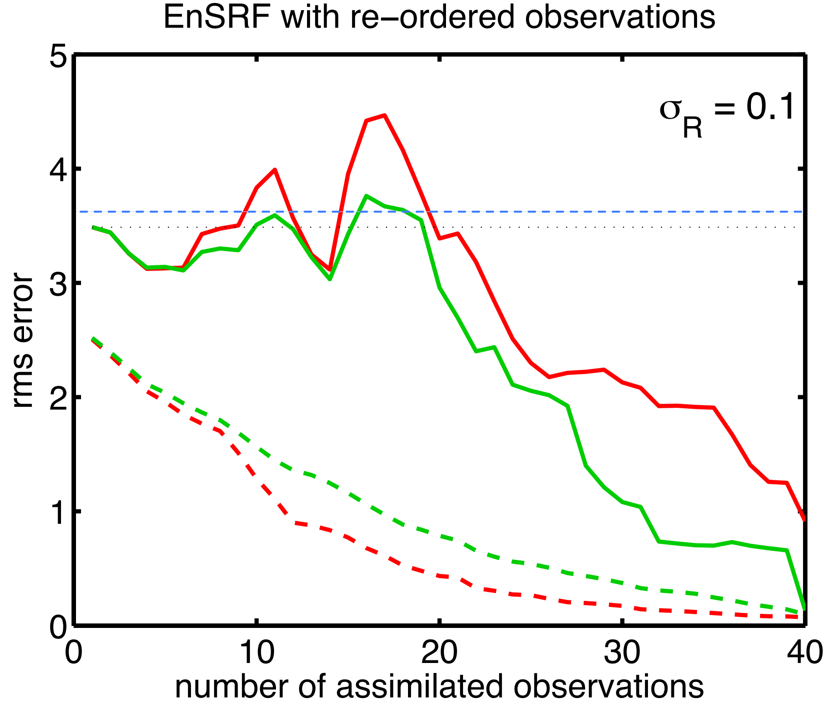
Blue: LESTKF with serial observation processing

lars.nerger@awi.de

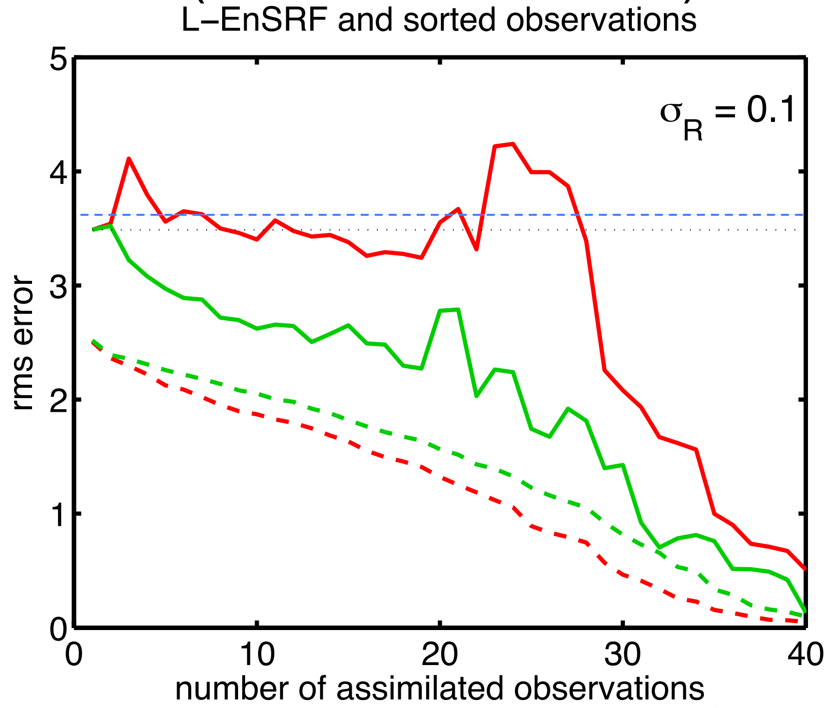
Effect of observation reordering

- Before: Assimilated observation from grid point 1 to 40 with increasing index
- What is the effect when we re-order the observations?

„maximum distance“
(1, 21, 11, 31, 6, 26,...)



local observation sorting
(Whitaker et al. 2008)



lars.nerger@awi.de

Serial obs. Processing and localization

- Update of covariance matrix is inconsistent
  - Because of asymmetric update equation
  - Because of small ensemble
  - Also the case for synchronous assimilation of observations (LETKF, LSEIK, LESTKF)

- Instability of serial observation processing
  - only significant when assimilation has strong influence (large state error and small observation error)
  - Can it happen, when ensemble spread gets large, e.g. due to ocean eddies, convection in atmosphere?

lars.nerger@awi.de

Summary

- Localization
  - is empirical
  - it works
  - regularizes the filter analysis step
  - does inconsistent covariance updates

- Optimal radius influenced by degrees of freedom from ensemble
- Interaction of localization and covariance matrix update still open

Thank you!

lars.nerger@awi.de